

Autonomous action robot for performing actions in response to human voice, has controller which controls direction of imaging unit, based on specified sound source direction and target image

Patent Assignee: HIGAKI N; HONDA GIKEN KOGYO KK; HONDA MOTOR CO LTD; SAKAGAMI Y; SUMIDA N; YOSHIDA Y

Inventors: HIGAKI N; SAKAGAMI Y; SUMIDA N; SUMITA N; YOSHIDA Y

Patent Family (3 patents, 2 countries)							
Patent Number	Kind	Date	Application Number	Kind	Date	Update	Type
US 20030055532	A1	20030320	US 2002225287	A	20020822	200339	B
JP 2003062777	A	20030305	JP 2001251642	A	20010822	200339	E
US 6853880	B2	20050208	US 2002225287	A	20020822	200511	E

Priority Application Number (Number Kind Date): JP 2001251642 A 20010822; US 2002225287 A 20020822

Patent Details					
Patent Number	Kind	Language	Pages	Drawings	Filing Notes
US 20030055532	A1	EN	18	10	
JP 2003062777	A	JA	10		

Alerting Abstract: US A1

NOVELTY - A sound source locating unit specifies the direction of sound source, based on the direction at which the sound is emitted. An extraction unit extracts target image of specific shape from an image picked up by an imaging unit. Based on the specified sound source direction and target image, a controller controls the direction of imaging unit.

USE - For performing actions in response to human voice.

ADVANTAGE - Recognizes face of person and performs various actions in response to commands.

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-062777

(43)Date of publication of application : 05.03.2003

(51)Int.Cl.

B25J 13/00

A63H 11/00

B25J 5/00

G06T 1/00

G06T 7/00

G06T 7/60

G10L 15/00

G10L 15/28

(21)Application number : 2001-251642

(71)Applicant : HONDA MOTOR CO LTD

(22)Date of filing : 22.08.2001

(72)Inventor : SAKAGAMI YOSHIKAKI

HIGAKI NOBUO

SUMITA NAOAKI

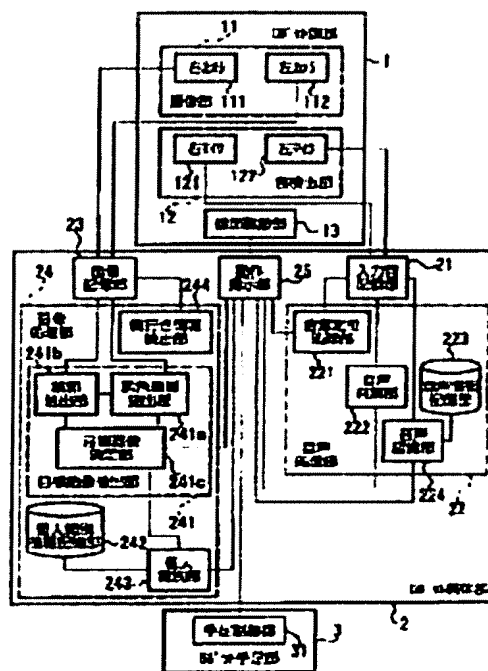
YOSHIDA YUICHI

(54) AUTONOMOUS ACTING ROBOT

(57)Abstract:

PROBLEM TO BE SOLVED: To provide an autonomous acting robot that can turn the line of sight to the direction of a calling person, recognize the face of a facing person and also take several actions corresponding to commands.

SOLUTION: At first a sound emitted from a person or other sound sources is detected by a sound detecting element 12. Then a sound source positioning processor 221 is operated based on the detected sound and the direction of the sound source where the sound is emitted is defined. Furthermore, by operating a head driving part 13, a robot head 1 equipping an imaging part 11 that shoots an image of periphery is controlled and then the shooting direction is directed to the direction of the defined sound source. And by operating the imaging part 11, the image of the periphery of the sound source direction is taken. Then a target image having a particular form is cut from the taken image by operating a target image cutting part 241. And by a command of an action command part 25, the imaging direction of robot head 1 is controlled and directed to the cut target image direction.



(11)特許出願公開番号

(P2003-62777A)

(43)公開日 平成15年3月5日(2003.3.5)

(51)Int.Cl. ⁷	識別記号	F I	テーマコード*(参考)
B 2 5 J 13/00		B 2 5 J 13/00	Z 2 C 1 5 0
A 6 3 H 11/00		A 6 3 H 11/00	Z 3 C 0 0 7
B 2 5 J 5/00		B 2 5 J 5/00	C 5 B 0 5 7
G 0 6 T 1/00	3 4 0	G 0 6 T 1/00	3 4 0 A 5 D 0 1 5
7/00	1 0 0	7/00	1 0 0 C 5 L 0 9 6
審査請求 未請求 請求項の数 6 O L (全 10 頁) 最終頁に続く			

(22) 出願日 平成13年8月22日(2001.8.22)

(71)出願人 000005326

本田技研工業株式会社

東京都港区南青山二丁目1番1号

(72) 発明者 坂上 義秋

埼玉県和光市中央1丁目4番1号 株式会社
本田技術研究所内

(72)発明者 桧垣 信男

埼玉県和光市中央1丁目4番1号 株式会社本田技術研究所内

(74) 代理人 100064908

弁理士 志賀 正武 (外5名)

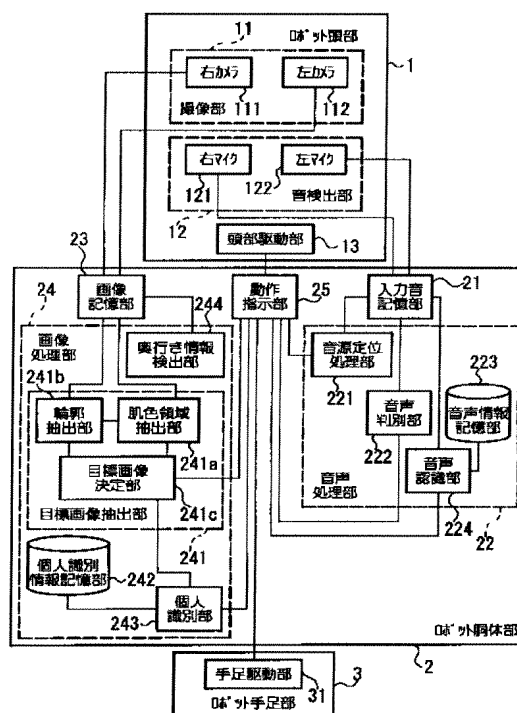
[最終頁に続く](#)

(54) 【発明の名称】 自律行動ロボット

(57) 【要約】

【課題】 視線を呼びかけた人の方向に向けたり、対面者の顔を認識したり、いくつかの命令に応じた行動をとることができる自律行動ロボットを提供する。

【解決手段】 まず、人物その他の音源から発せられた音が、音検出部１２により検出される。次に、検出された音に基づいて、音源定位処理部２２１が作動することにより、その音を発した音源の方向が特定される。さらに、頭部駆動部１３の作動により、周辺の画像を撮影する撮像部１１が設置されたロボット頭部１が制御されて、その撮影方向が特定された音源の方向に向けられる。そして、撮像部１１の作動により、音源の方向の周辺の画像が撮影される。次いで、目標画像抽出部２４１の作動により、撮影された画像から特定の形状を有する目標画像が抽出される。そして、動作指示部２５の指示により、ロボット頭部１の撮影方向が制御されて抽出された目標画像の方向に向けられる。



【特許請求の範囲】

【請求項1】 音源から発せられた音を検出する音検出手段と、
 検出された音に基づいて音源の方向を特定する音源定位手段と、
 周辺の画像を撮影する撮像手段と、
 撮影された画像から特定の形状を有する目標画像を抽出する目標画像抽出手段と、
 特定された音源方向および抽出された目標画像に基づいて該目標画像に向かう方向に前記撮像手段の撮影方向を制御する制御手段とを有することを特徴とする自律行動ロボット。

【請求項2】 前記特定の形状が人間の顔の形状であり、
 特定の人間の顔を識別する識別手段をさらに備えることを特徴とする請求項1記載の自律行動ロボット。

【請求項3】 検出された音が特定の音声命令であることを認識する音声認識手段と、
 認識された音声命令に基づいて姿勢または位置を変更する姿勢位置変更手段とをさらに備えることを特徴とする請求項1または2記載の自律行動ロボット。

【請求項4】 音源から発せられた音を検出する音検出手段と、
 検出された音に基づいて音源の方向を特定する音源定位手段と、
 周辺の画像を撮影する撮像手段と、
 撮影された画像から特定の形状を有する目標画像を抽出する目標画像抽出手段と、
 特定された音源方向および抽出された目標画像に基づいて該目標画像に向かう方向に姿勢または位置を変更する姿勢位置変更手段とを備えることを特徴とする自律行動ロボット。

【請求項5】 撮影された画像に動的輪郭モデルを適用することによって前記目標画像の輪郭を抽出する輪郭抽出手段をさらに備えることを特徴とする請求項1から4までのいずれかの項に記載の自律行動ロボット。

【請求項6】 前記撮像手段が、互いに対応付けられた2台の撮像装置から構成され、
 撮影された2枚の画像で構成されるステレオ画像を立体視することによって奥行き情報を検出する奥行き情報検出手段をさらに備えることを特徴とする請求項1から5までのいずれかの項に記載の自律行動ロボット。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、画像処理と音声処理とを併用して移動や姿勢制御を行う自律行動ロボットに関する。

【0002】

【従来の技術】近年、人間のパートナーとして共存することを目的とした自律行動ロボットが開発されてきてい

る。この種の自律行動ロボットには、人間の音声に反応してあらかじめ決められた行動をするものがある。また、カメラ等の各種センサを用いて、認識対象である目標を含んだシーンをイメージ情報として取得し、そのイメージ情報と目標に対する断片的な知識とを利用して目標を認識するものもある。例えば、人間の音声を認識し、音声の種類によっていくつかの動作によってできるだけ人間らしい振る舞いをさせ、親近感を感じさせるような人間型ロボットも自律行動ロボットの一つの実施形態である。

【0003】

【発明が解決しようとする課題】しかしながら、本物の人間により近い動作をする人間型ロボットを開発するためには、例えば、ロボットの視野外にいる対面者の呼びかけに反応して振り向いたり、対面者の顔を見分けたりするような機能を備える必要がある。さらに、対面者がジェスチャー等によりある物を提示したような場合には、指示された物等に視線を移して、その物を認識することができるよう機能がこれからの人間型ロボットには必要である。

【0004】この発明は、このような事情を考慮してなされたものであり、視線を呼びかけた人の方向に向けたり、対面者の顔を認識したり、いくつかの命令に応じた行動をとることができるような自律行動ロボットを提供することを目的とする。

【0005】

【課題を解決するための手段】上記課題を解決するために、請求項1に記載の発明は、音源から発せられた音を検出する音検出手段と、検出された音に基づいて音源の方向を特定する音源定位手段と、周辺の画像を撮影する撮像手段と、撮影された画像から特定の形状を有する目標画像を抽出する目標画像抽出手段と、特定された音源方向および抽出された目標画像に基づいて該目標画像に向かう方向に前記撮像手段の撮影方向を制御する制御手段とを有することを特徴とする。

【0006】この発明によれば、まず、人物その他の音源から発せられた音が、音検出手段により検出される。次に、検出された音に基づいて、音源定位手段が作動することにより、その音を発した音源の方向が特定される。さらに、制御手段の作動により、周辺の画像を撮影する撮像手段であるカメラが制御されて、その撮影方向が特定された音源の方向に向けられる。そして、撮像手段の作動により、音源の方向の周辺の画像が撮影される。次いで、目標画像抽出手段の作動により、撮影された画像から特定の形状を有する目標画像が抽出される。そして、制御手段の作動により、撮影手段の撮影方向が制御されて抽出された目標画像の方向に向けられる。したがって、自律行動ロボットに搭載されたカメラの視野外にいる人物等が音を発した場合でも、その音に反応して当該人物等の方向にカメラの撮影方向を自動的に向け

させることが可能となる。

【0007】請求項2記載の発明は、請求項1記載の発明において、前記特定の形状が人間の顔の形状であり、特定の人間の顔を識別する識別手段をさらに備えることを特徴とする。

【0008】この発明によれば、目標画像抽出手段の作動により、撮影された画像から人間の顔の形状を有する目標画像が抽出される。そして、抽出された目標画像に基づいて、識別手段の作動により、特定の人間の顔が識別される。したがって、音声を発した人物の顔を抽出して識別することにより、特定の人物の個人認証を行うことが可能となる。

【0009】請求項3記載の発明は、請求項1または2記載の発明において、検出された音が特定の音声命令であることを認識する音声認識手段と、認識された音声命令に基づいて姿勢または位置を変更する姿勢位置変更手段とをさらに備えることを特徴とする。

【0010】この発明によれば、音検出手段の作動により検出された音が特定の音声命令であることが、音声認識手段の作動により認識される。そして、認識された音声命令に基づいて、姿勢位置変更手段の作動により、自律行動ロボット自身の姿勢または位置が変更される。したがって、人間が発した音声命令の種類によって、あらかじめ決められた姿勢変化あるいは移動動作を行うことが可能となる。

【0011】請求項4記載の発明は、音源から発せられた音を検出する音検出手段と、検出された音に基づいて音源の方向を特定する音源定位手段と、周辺の画像を撮影する撮像手段と、撮影された画像から特定の形状を有する目標画像を抽出する目標画像抽出手段と、特定された音源方向および抽出された目標画像に基づいて該目標画像に向かう方向に姿勢または位置を変更する姿勢位置変更手段とを備えることを特徴とする。

【0012】この発明によれば、まず、人物その他の音源から発せられた音が、音検出手段の作動により検出される。次に、検出された音に基づいて、音源定位手段が作動することにより、その音を発した音源の方向が特定される。一方、撮像手段の作動により、自律移動ロボット周辺の画像が撮影される。次いで、目標画像抽出手段の作動により、撮影された画像から特定の形状を有する目標画像が抽出される。そして、姿勢位置変更手段の作動により、特定された音源方向および抽出された目標画像に基づいて目標画像に向かう方向に姿勢または位置が変更される。したがって、人物の位置を発生された音声により認識するだけでなく、人間のような振る舞いをして、その人物に対して親近感を抱かせることが可能となる。

【0013】請求項5記載の発明は、請求項1から4までのいずれかの項に記載の発明において、撮影された画像に動的輪郭モデルを適用することによって前記目標画

像の輪郭を抽出する輪郭抽出手段をさらに備えることを特徴とする。

【0014】この発明によれば、画像入力手段の作動により入力された画像にスネーク（Snakes）と呼ばれる動的輪郭モデルを適用する輪郭抽出手段の作動により、撮影された画像の中から目標画像の輪郭が抽出される。したがって、動的輪郭モデルを用いた移動体抽出アルゴリズムによって、人物の輪郭を高精度で抽出することが可能となる。

10 【0015】請求項6記載の発明は、請求項1から5までのいずれかの項に記載の発明において、前記撮像手段が、互いに対応付けられた2台の撮像装置から構成され、撮影された2枚の画像で構成されるステレオ画像を立体視することによって奥行き情報を検出する奥行き情報検出手段をさらに備えることを特徴とする。

【0016】この発明によれば、人物等の画像を撮影する撮像手段が2台のカメラである撮像装置から構成され、その2台のカメラによってステレオ画像が撮影される。そして、奥行き情報検出手段の作動により、撮影されたステレオ画像を立体視することによって、カメラからの距離を示す奥行き情報が検出される。したがって、音を発した人物等の方向だけではなく、カメラとの3次元位置関係に関する情報を検出することが可能となる。

【0017】

【発明の実施の形態】以下、図面を参照して、この発明の一実施形態について説明する。図1は、この発明の一実施形態による自律行動ロボットの構成を示すブロック図である。本実施形態における自律行動ロボットは、対面者である人物に対して親近感を抱かせるような人間型ロボットを想定しており、目や耳等を有する頭部に相当するロボット頭部1と、胴体に相当するロボット胴体部2と、手足に相当するロボット手足部3とから構成される。尚、犬や猫のタイプのような4本足のペットロボットの場合、ロボットの手や足は4つの部分に分かれてそれぞれの部分が関連して動作するが、本実施形態では便宜上これをまとめて一つのロボット手足部3として記述する。

30

【0018】ロボット頭部1は、周囲のカラー画像を撮影するための撮像部11と、音源から発せられた音を検知するための音検出部12と、撮像部11および音検出部12が固定されているロボット頭部1自身を左右上下方向に任意の角度で向けることができる頭部駆動部13とから構成される。

40

【0019】撮像部11は人間型ロボットの目の役割を果たすため、ロボット頭部前面に同じ高さに水平に設置された、互いに対応付けられた2台のCCDカメラで構成されており、カラーのステレオ画像を撮影することが可能である。これらのカメラは図1において右カメラ111と左カメラ112で示されている。尚、本実施形態では入力画像は8ビット（256階調）の濃淡値で表現

50

されているものとする。また、音検出部12は、人間型ロボットの耳の役割を果たすため、ロボット頭部1の左右それぞれの側面に設置された2本のマイクで構成されている。図1において、その2本のマイクは、右マイク121と左マイク122で示されている。これらのマイクで受信したアナログの入力音響信号は、デジタルの時系列信号に変換される。

【0020】ロボット胴体部2は、撮像部11で撮影されたカラーのステレオ画像を記憶するための画像記憶部23と、記憶された画像を処理してロボット頭部1またはロボット手足部3にあらかじめ決められた動作をさせるための情報を獲得する画像処理部24とを備える。また、ロボット胴体部2は、音検出部12で検出された音に関する情報を記憶する入力音記憶部21と、入力された音を処理してロボット頭部1にあらかじめ決められた動作をさせるための情報を獲得する音声処理部22とを備える。さらにロボット胴体部2は、音声処理部22と画像処理部24における処理の結果に基づいて、ロボット頭部1またはロボット手足部3の動作を制御して、ロボット頭部1の向きを変化させ、ロボットの姿勢や位置を変化させるための動作指示を行う動作指示部25を備えている。

【0021】音声処理部22は、記憶された音に関する情報に基づいて両マイクの音到達時間差と音圧差により水平方向の音源の相対的な方向を特定する音源定位処理部221と、当該音が人間が発した音声か否かを判別するための音声判別部222と、特定の音声命令が記憶された音声情報記憶部223と、音声情報記憶部223に接続され入力された音声が特定の命令であるか否かを認識する音声認識部224とから構成される。

【0022】一方、画像処理部24は撮影された画像から特定の目標を抽出するための目標画像抽出部241と、特定の個人を識別するための個人情報記憶された個人識別情報記憶部242と、目標画像抽出部241と個人識別情報記憶部242とに接続され、抽出された人物の顔画像が特定の個人であるか否かを識別する個人識別部243と、画像記憶部23に記憶されたステレオ画像から立体視によって奥行き情報を算出する奥行き情報検出部244とから構成される。

【0023】また、目標画像抽出部241は、人物の顔画像等を含む肌色領域を抽出するための肌色領域抽出部241aと、人物等の輪郭を抽出するための輪郭抽出部241bと、肌色領域抽出部241aまたは輪郭抽出部241bで抽出された情報に基づいて特定の目標画像を決定する目標画像決定部241cとから構成される。まず、肌色領域抽出部241aは、RGB座標で取得したカラー画像をHLS座標に色座標変換して、色相成分等を用いて肌色に相当する領域だけを抽出することによって肌色領域からなる肌色領域画像を作成する。また、輪郭抽出部241bは、肌色領域抽出手段241aで抽出

された肌色領域の輪郭部分をエッジとして抽出してエッジ画像を作成する手段である。尚、肌色領域抽出部241aと輪郭抽出部241bにおける処理は、特願2001-013818に記載されている肌色領域抽出及び輪郭抽出の方法を用いることもできる。

【0024】また、輪郭抽出部241bは上記抽出方法に加えて、スネーク(Snakes)と呼ばれる動的輪郭モデルを適用する輪郭抽出手段の作動により、撮影された画像の中から目標画像の輪郭を抽出することもできる。さらに、ロボット手足部3には、ロボット胴体部2の動作指示部25に接続され、当該ロボットの手足をあらかじめ決められた動作を行うように制御するための手足駆動部31が備わっている。このように、上述した頭部駆動部13と動作指示部25と手足駆動部31は、人間型ロボットの動作等の制御手段となり対面者に対して親しみを抱かせるような動きをする。

【0025】次に、図面を参照して、上述した実施形態の動作について説明する。図2は、同実施形態における自律行動ロボットの各部の動作を説明するためのフローチャートである。以下の説明において、画像処理、音源定位処理、音声認識処理、動作指示処理は各々並列に処理が実行され、それらの処理結果に基づいて視線制御値の算出を行い頭部の駆動が行なわれるものとする。まず、音源方向、音種別、音量を入力する(ステップS1)。続いて、音声認識結果を入力する(ステップS2)。さらに、頭部位置候補、及び顔位置候補を入力する(ステップS3)。

【0026】そして、音の入力があったか否かが判断される(ステップS4)。その結果、音検出部12で音が検出されなかった場合(NO)、撮影された画像に対して画像処理部24の作動により、目標画像の抽出処理が行われる(ステップS11)。また、何らかの音が検出された場合(YES)、その音に関する情報が入力音記憶部21に記憶される。ここで、音に関する情報とは、検出された音の受信信号や、左右マイク121、122に到達したときの時間、時間差等が含まれる。そして、これらの音声情報を音声処理部22で処理することにより動作指示部25に与える各種情報が求められる。

【0027】音声処理部22における音源定位処理部221では、当該ロボットに対する左右マイク121、122で検出された音の音源の方向が特定される(ステップS5)。ここで、音源の方向を特定する手順について図面を用いて詳述する。図3は、同実施形態による自律行動ロボットの音源定位処理部221の細部の構造を説明するためのブロック図である。図3において、音到達時間差計算部221aには、入力音記憶部21に記憶された入力音に関する情報に基づいて、左右マイク121、122で検出された音の到達時間が入力され、その時間差が計算される。すなわち、両入力時系列の相互相関関数から時系列間の位相差を推定し、音到達時間差 τ

の候補を出力する。

【0028】また、音圧測定部221b、cでは、それぞれ右マイク121、左マイク122で検出された音に関する受信信号から、入力時系列の一定時間窓における二乗平均の平方根を計算することによってdB値によって音圧 V_s が計算される。そして、音圧差計算部221dにおいて音圧計算部221bおよび221cで求められた音圧の差がdB値で計算される。その結果、音源方向計算部221eにおいて、右マイク121と左マイク122で検出された音から音源の方向 θ_s を出力することができる。

【0029】ここで、音圧の計算方法について説明する。図4は、入力初期位置の音圧の時間変化から推定されたエンベロープ（波形形状）による音圧の計算例について示した図である。例えば、図4（a）に示すように、人間の音声「おい」が入力され、左図のように振幅と時間との関係が得られた場合、エンベロープ推定によって右図のように音圧と時間との関係に置き換えることができる。そこで、エンベロープのなだらかさを評価して、なだらかな場合を人間の音声と判定して V_s の符号を正の値とし、急峻な立ち上がりをする場合を音声以外の音として V_s の符号を負に設定する。また、図4

（b）に示すように、手拍子音の場合は、人間の音声と比較して、振幅は音声よりも大きい、計測された時間が短いという特徴がある。

【0030】一方、図5は、図3に示された音源定位処理部221において出力である音源方向を求めるための手順について説明するための図である。ここでは、2つの音響入力端子（右マイク121、左マイク122）に到達する音響信号における時間差と音圧差により、水平方向の相対的な音源の方向を推定する。図5（a）は、音源定位処理部221における動作手順について詳細に説明するための図である。図3に示す音到達時間差計算部221aにおいて、音到達時間差 τ の候補が計算される。また、音圧差計算部221dにおいて、音圧差 V_s が計算される。次に、図5（b）に示すような音圧差からの音到達時間差 τ の選択領域マップを用いて、求めた音圧差から音到達時間差 τ の選択領域Aを設定する。そして、音到達時間差 τ の候補から選択領域Aに属する τ を選択する。このようにして求めた音到達時間差 τ から、以下に示す音源の方向を算出する計算式、あるいは図4（c）に示される τ と θ_s とを対応付けるマップによって音源の方向 θ_s を求める。

【0031】図6は、本実施形態における音源の方向 θ_s を説明するための図である。図6に示すように、音源の方向は、左右マイク121、122を結んだ線分と、その線分の中点と音源とを結んだ線分とのなす角 θ_s で示す。すなわち、音源の方向は式（1）に示される角度 θ_s で求めることができる。

$$\theta_s = \sin^{-1} (V \cdot \tau / d) \quad (1)$$

ここで、 V は音速、 τ はそれぞれのマイクで検出された音の到達時間差、 d はそれぞれのマイク間の距離である。

【0032】これと同時に、検出された音が人間の音声であるか、それ以外の音であるかの判別が音声判別部222において行われる（ステップS5）。図7は、音声判別部222の細部構成を説明するためのブロック図である。ここでは、音響入力端子（右マイク121、左マイク122）に到達する音響信号の入力初期位置におけるエンベロープ（波形形状）を評価することにより、その音響信号が音声であるか、それ以外の音であるかを判別する。図7において、音圧計算部222aと222bは、それぞれ右マイク121と左マイク122で検出した音の音圧を計算する部分である。また、平均音圧計算部222cは、計算した音圧の平均音圧を計算する部分である。さらに、エンベロープ推定部222dは、図4に示す振幅からエンベロープを推定する部分である。そして、判別部222eは、人間の音声であるか、それ以外の音であるかを判別する部分である。

【0033】図8は、図7に示される音声判別部222の動作を説明するためのフローチャートである。入力された音は上述したように、音圧が計算されて平均音圧計算部222cの作動により、平均音圧 V_s が計算され、エンベロープ推定部222dで図4に示すようなエンベロープが推定される（ステップS81）。そして、計算された平均音圧 V_s があらかじめ設定された一定値 α 以上であるか否かが判断される（ステップS82）。その結果、平均音圧 V_s が一定値 α 以上であれば（YES）、続いて平均音圧 V_s と音響信号先頭の音圧 $p V_s$ とを比較して β 以上減衰しているか否かが判断される（ステップS83）。一方、ステップS82で平均音圧 V_s が一定値 α 以上でない場合（NO）、人間の音声以外の音であると判断される（ステップS84）。

【0034】そして、ステップS83において、 β 以上減衰していると判断された場合（YES）、人間の音声以外の音であると判断される（ステップS84）。また、 β 以上減衰していないと判断された場合（NO）、ステップS85に進んで一定時間 T 以上継続した音か否かが判断される。そして、一定時間 T 以上の音であると判断された場合（YES）、人間の音声であると判断される（ステップS86）。また、一定時間 T 以上継続していない場合（NO）、ステップS81に戻る。上述したステップS82以降の一連の処理は、判別部222eの作動により行われる。尚、音圧計算部は音源定位処理部221と音声判別部222の両方に必要であるため、音源定位処理部221と音声判別部222の両部で音圧計算部を共有する構造にしてもよい。

【0035】次に、このようにして求められた音源の位置が人間型ロボットの視野角内にあるか否かが判定される（ステップS6）。その結果、すでにロボットの視野

角内にあれば、ステップS8に進む。また、視野角内ないと判断された場合は、ロボット頭部1の方向が求められた音源の方向 θ_s に基づいて動作指示部25からの指示による頭部駆動部13の作動により方向転換される(ステップS7)。そして、再びステップS1へ戻る。

【0036】さらに、ステップS5で判定された入力音が人間の音声であるか否かが判定される(ステップS8)。その結果、人間の音声であると判断された場合(YES)、さらにその音声が指示語であるか否かが判断される(ステップS10)。また、ステップS8で人間の音声でないと判断された場合(NO)、一定時間経過後に方向転換したロボット頭部1を元の方向に戻す(ステップS9)。

【0037】ステップS10では、入力された音声があらかじめ決められた指示語であるか否かが判定される。この判定は、音声情報記憶部223に記憶されている音声に関する情報と入力音声との比較が音声認識部224において行われる。この結果、入力音声が特定の指示語であると判定された場合は、指示された手先位置の領域が推定される(ステップS14)。次いで、この方向に基づいて視線制御値が算出される(ステップS15)。そして、ロボット頭部1が方向転換される(ステップS16)。

【0038】一方、画像処理では人間の顔領域の抽出が行われる(ステップS11)。顔領域の抽出に際して、肌色領域および輪郭の抽出方法については、上述したように、肌色領域抽出部241aは、RGB座標で取得したカラー画像をHLS座標に色座標変換して、色相成分等を用いて肌色に相当する領域だけを抽出することによ

$$G(x, y) = 1/n * \{M(x, y) + F(x, y)\} \quad (2)$$

但し、nは互いの距離がしきい値内の位置の総数である。

【0040】そして、この中から移動体までの距離が最も近い $G(x, y)$ を選んで、これに基づいてパン、チルトの視線制御値を算出することによって、ロボット頭部1が制御される。尚、移動体までの距離の選択方法は、移動体が複数の場合は、適切な順によってロボット頭部1のパン、チルト制御させ、くまなくロボットの視線を制御させることも可能である。また、音の入力がなく、顔候補位置 $F(x, y)$ のみが求められた場合には、画像中央からの距離が近い順にパン、チルト角を算出して、ロボット頭部1を制御させる。さらに、音の入力がなく、頭部候補位置 $M(x, y)$ のみが求められた場合には、距離が最も近いものから順にパン、チルト角を算出して、ロボット頭部1を制御させる。

【0041】一方、ステップS10で入力音声が、例えば「これ」、「この」、「ここ」等の指示語であった場合、手先位置 $H(x, y)$ が推定される(ステップS14)。すなわち、図9に示すように、頭部候補位置 $M(x, y)$ と輪郭内画素の位置の平均から求められる移

って肌色領域とそれ以外の領域とからなる肌色領域画像を生成する。また、輪郭抽出部241bは、肌色領域抽出部241aにおいて得られた肌色領域画像からエッジ抽出をすることによって肌色領域の輪郭を抽出する。尚、顔領域の抽出において、画像内輪郭の最上点を始点として、あらかじめ距離に応じた大きさの矩形領域を走査して、輪郭内画素が最も多く含まれる位置を求めて、その位置を頭部候補位置 $M(x, y)$ とする。ここで、 (x, y) は原点を画像の左上として、右方向をx、左方向をyとした画像上の座標値を表している。図9は、移動体の輪郭と頭部候補位置および手先位置を説明するための図である。尚、図9(c)に示す手先位置に関する処理については後述する。図10は、撮影されたカラー画像と抽出される顔画像候補 $F(x, y)$ との関係を示明するための図である。

【0039】そして、上述した手順によって得られた視線制御値である音源の方向 θ_s と、音声認識結果 C_v と、頭部候補位置 $M(x, y)$ と、顔候補位置 $F(x, y)$ とが算出される(ステップS12)。これらのデータが動作指示部25に入力することによって、頭部駆動部13あるいは手足駆動部31の動作が制御され、ロボット頭部1あるいはロボット手足部3が動く(ステップS13)。例えば、頭部候補位置 $M(x, y)$ または顔候補位置 $F(x, y)$ の相対角と音源の方位 θ_s との差が最小となる頭部候補位置 $M(x, y)$ または顔候補位置 $F(x, y)$ を選択して、それよりパン、チルト角が算出される。但し、音の入力がない場合は、頭部候補位置 $M(x, y)$ および顔候補位置 $F(x, y)$ 間の距離が、あるしきい値内であるものについて式(2)で求める。

動体重心Gからより遠い輪郭上の位置であって、かつ、重心Gより上の部分にあたる点 $H(x, y)$ が手先位置として目標画像推定部241cにおいて推定される。ここで、 $H(X, Y)$ を中心とする矩形領域内から肌色領域を抽出し、手先位置を求めてもよい。次いで、その位置に基づいて、頭部駆動部13に対してパン、チルト角等の動作指示情報が動作指示部25において算出される(ステップS15)。そして、動作命令が頭部駆動部13に伝えられて、例えばロボット頭部1が方向転換される(ステップS16)。

【0042】また、本実施形態においては、輪郭抽出手段として上述した差分による方法のほかに、ステレオ画像から得られる時系列画像からスネークと呼ばれる移動体抽出アルゴリズムを用いて移動体の輪郭を抽出してもよい。この方法においては、ある探索領域内において動きのあるエッジが最も多く分布する距離を移動体までの距離と仮定して、その距離近辺にあるエッジを画像エネルギーとする。そして、この画像エネルギーを外部エネルギーとしてスネークアルゴリズムを用いて輪郭を抽出する。

【0043】なお、図1における音声処理部22および画像処理部24の機能を実現するためのプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することにより顔領域の検出等を行ってもよい。なお、ここでいう「コンピュータシステム」とは、OSや周辺機器等のハードウェアを含むものとする。また、「コンピュータシステム」は、WWWシステムを利用している場合であれば、ホームページ提供環境（あるいは表示環境）も含むものとする。

【0044】また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、CD-ROM等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムが送信された場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリ（RAM）のように、一定時間プログラムを保持しているものも含むものとする。

【0045】また、上記プログラムは、このプログラムを記憶装置等に格納したコンピュータシステムから、伝送媒体を介して、あるいは、伝送媒体中の伝送波により他のコンピュータシステムに伝送されてもよい。ここで、プログラムを伝送する「伝送媒体」は、インターネット等のネットワーク（通信網）や電話回線等の通信回線（通信線）のように情報を伝送する機能を有する媒体のことをいう。また、上記プログラムは、前述した機能の一部を実現するためのものであっても良い。さらに、前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるもの、いわゆる差分ファイル（差分プログラム）であっても良い。

【0046】

【発明の効果】以上説明したように、請求項1記載の発明によれば、自律行動ロボットに搭載されたカメラの視野外にいる人物等が音を発した場合でも、その音に反応して当該人物等の方向にカメラの撮影方向を自動的に向けさせることができる。また、請求項2記載の発明によれば、音声を発した人物の顔を抽出して識別することにより、特定の人物の個人認証を行うことができる。さらに、請求項3記載の発明によれば、人間が発した音声命令の種類によって、あらかじめ決められた姿勢変化あるいは移動動作を行うことができる。

【0047】さらにまた、請求項4記載の発明によれば、人物の位置を発生された音声により認識するだけで

はなく、人間のような振る舞いをして、その人物に対して親近感を抱かせることができる。さらにまた、請求項5記載の発明によれば、動的輪郭モデルを用いた移動体抽出アルゴリズムによって、人物の輪郭を高精度で抽出することができる。さらにまた、請求項6記載の発明によれば、音を発した人物等の方向だけではなく、カメラとの3次元位置関係に関する情報を検出することができる。したがって、視線を呼びかけた人の方向に向けたり、対面者の顔を認識したり、いくつかの命令に応じた行動をとることができる。

【0048】

【図面の簡単な説明】

【図1】 この発明の一実施形態による自律行動ロボットの構成を示すブロック図である。

【図2】 同実施形態における自律行動ロボットの各部の動作を説明するためのフローチャートである。

【図3】 同実施形態による自律行動ロボットの音源定位処理部221の細部の構造を説明するためのブロック図である。

20 【図4】 入力初期位置の音圧の時間変化から推定されたエンベロープ（波形形状）による音圧の計算例について示した図である。

【図5】 図3に示された音源定位処理部221において出力である音源方向を求めるための手順について説明するための図である。

【図6】 本実施形態における音源の方向 θ_s を説明するための図である。

【図7】 音声判別部222の細部構成を説明するためのブロック図である。

30 【図8】 図7に示される音声判別部222の動作を説明するためのフローチャートである。

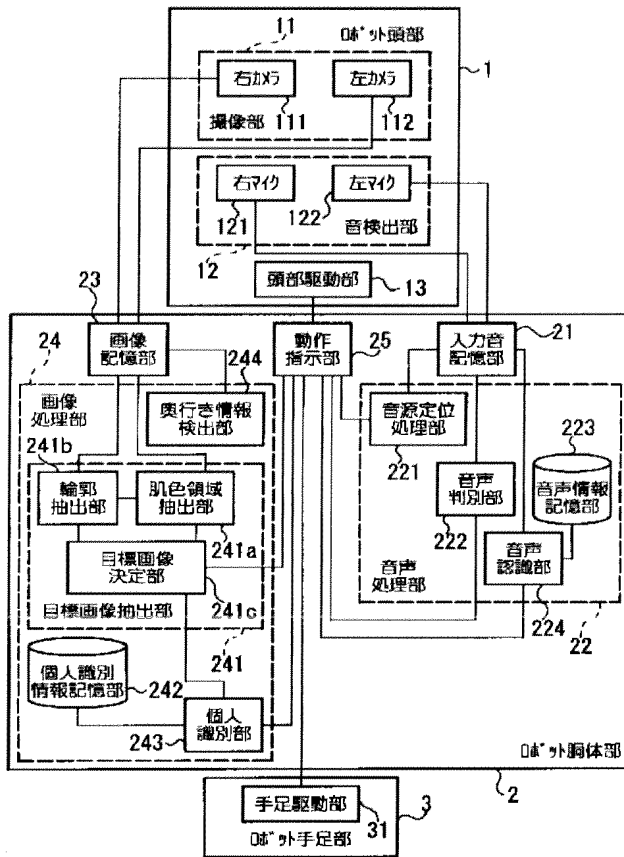
【図9】 移動体の輪郭と頭部候補位置および手先位置を説明するための図である。

【図10】 撮影されたカラー画像と抽出される顔画像候補M(x, y)との関係を説明するための図である。

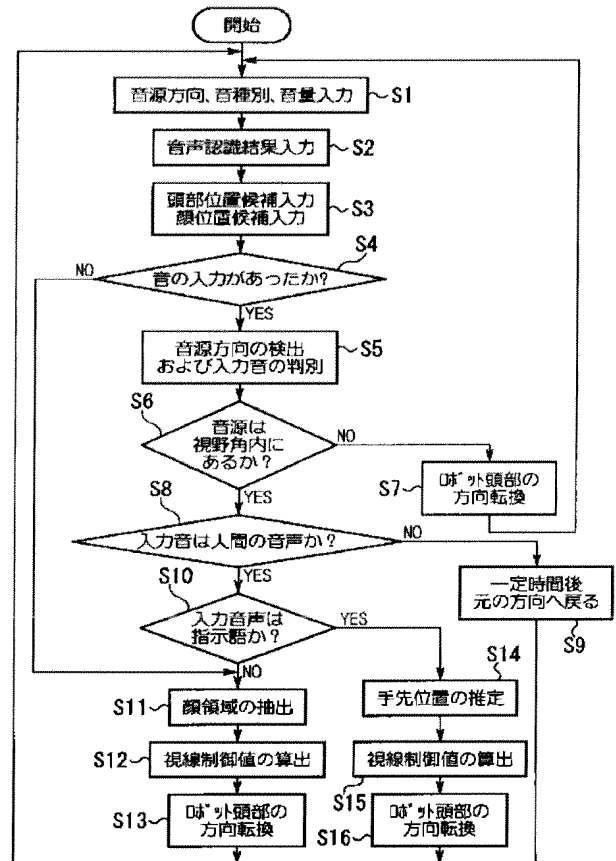
【符号の説明】

1 ロボット頭部、 2 ロボット胴体部、 3 ロボット手足部、 11 撮像部、 12 音検出部、 13 頭部駆動部、 22 音声処理部、 24 画像処理部、 25 動作指示部、 31 手足駆動部、 111 右カメラ、 112 左カメラ、 121 右マイク、 122 左マイク、 221 音源定位処理部、 222 音声判別部、 223 音声情報記憶部、 224 音声認識部、 241 目標画像抽出部、 242 個人識別情報記憶部、 243 個人識別部、 244 実行情報検出部

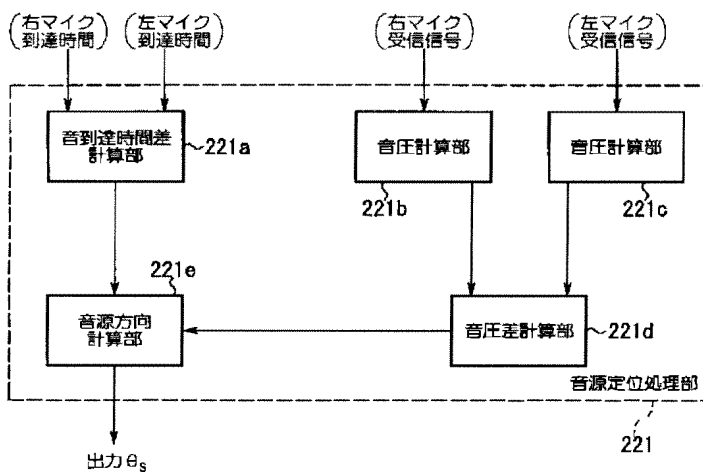
【图 1】



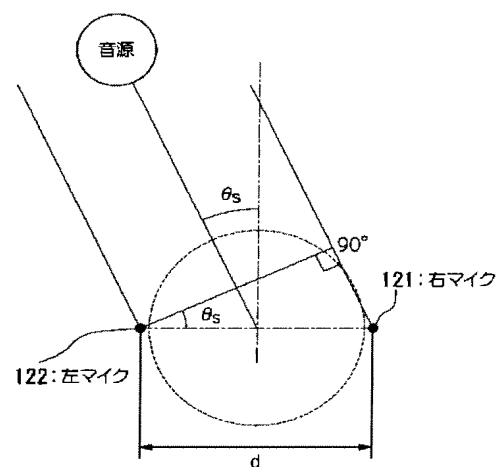
【図 2】



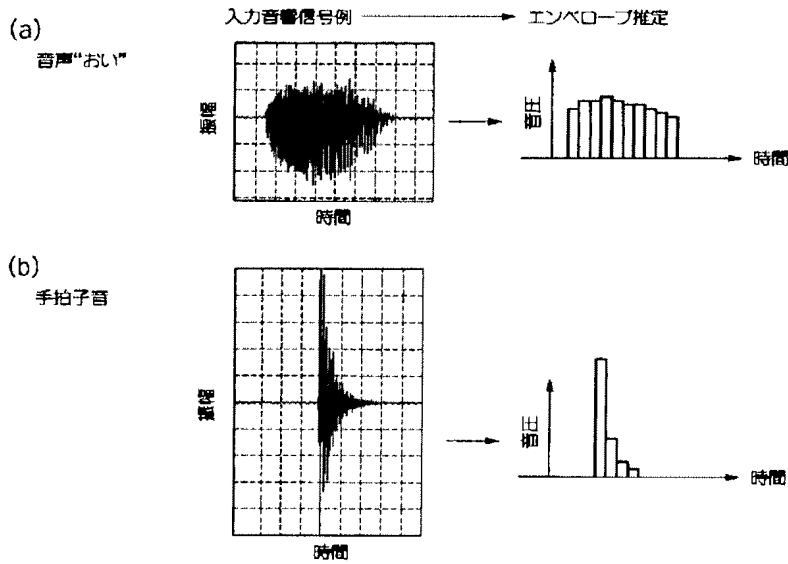
【图 3】



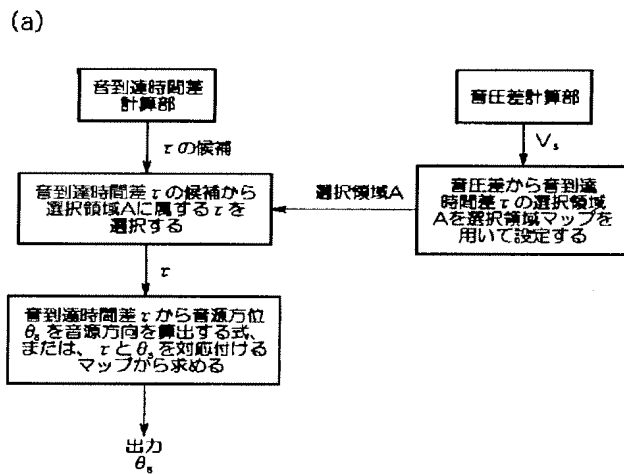
【図 6】



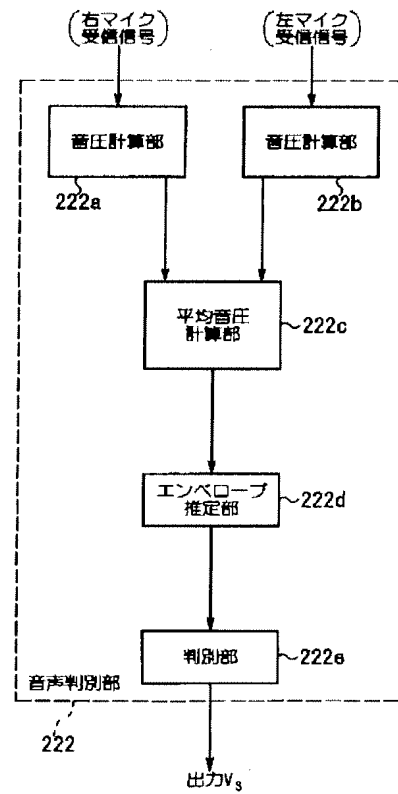
【図4】



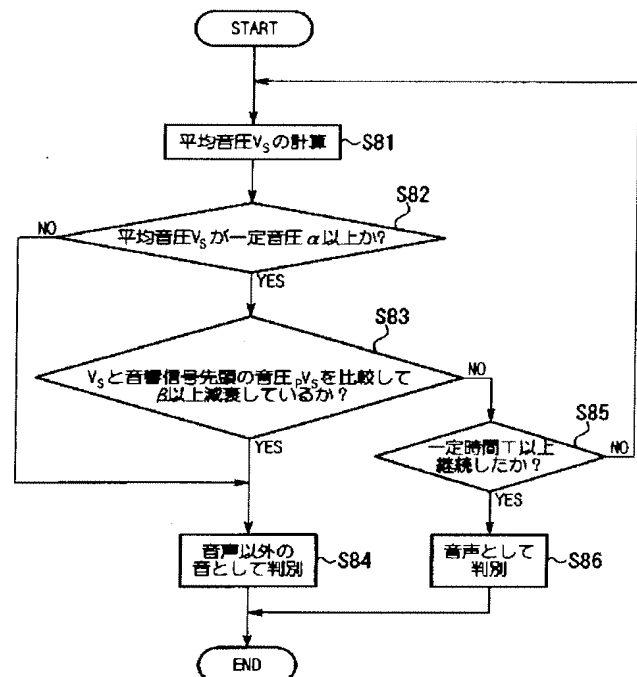
【図5】



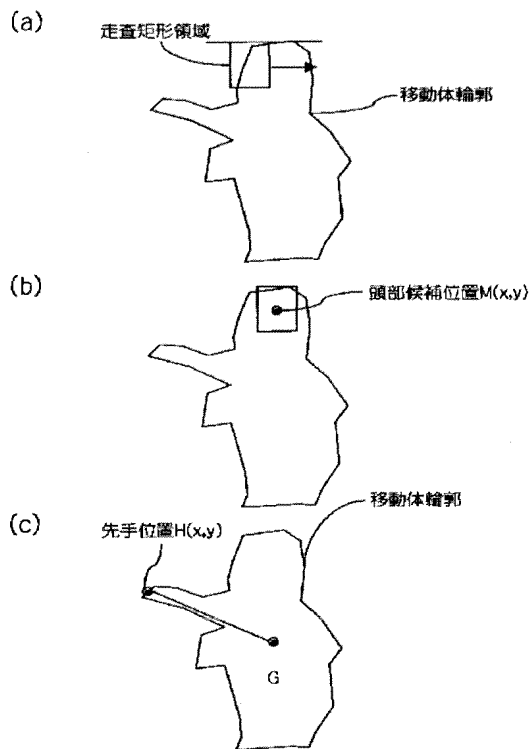
【図7】



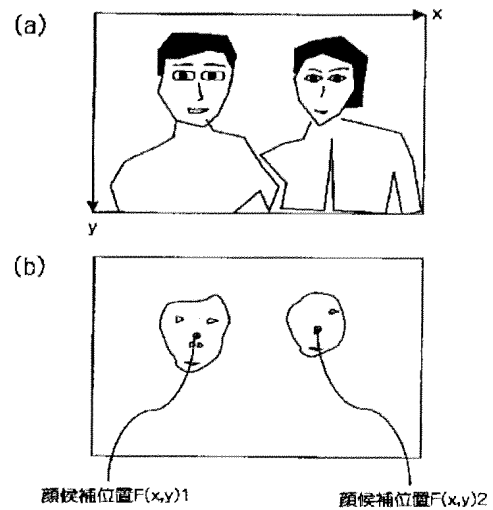
【図8】



【図 9】



【図 10】



フロントページの続き

(51)Int.Cl. ⁷	識別記号	F I	ターマコード [*] (参考)
G 0 6 T 7/60	2 5 0	G 0 6 T 7/60	2 5 0 C
G 1 0 L 15/00		G 1 0 L 3/00	5 5 1 H
15/28			5 1 1

(72)発明者 住田 直亮
 埼玉県和光市中央1丁目4番1号 株式会
 社本田技術研究所内
 (72)発明者 吉田 雄一
 埼玉県和光市中央1丁目4番1号 株式会
 社本田技術研究所内

F ターム(参考) 2C150 CA02 CA04 DA05 DA24 DA26
 DA27 DA28 DF03 DF04 DF06
 ED42 ED47 ED52 EF07 EF13
 EF16 EF17 EF23 EF28 EF29
 EF33 EF36
 3C007 AS36 BS27 CS08 KS04 KS11
 KS39 KT03 KT04 KT11 KX02
 MT14 WA03 WA04 WA13 WA14
 WB17 WB19
 5B057 BA11 BA13 CA01 CA08 CA12
 CA13 CA16 DA08 DB02 DB06
 DB09 DC09 DC16 DC33
 5D015 KK00
 5L096 AA02 AA06 BA05 CA04 CA05
 FA06 FA15 GA38 GA40 HA02